**Optimal Trading Strategies**  
in a Limit Order Market with Imperfect Liquidity

P. Kovaleva\(^a\)*, G. Iori\(^b\)

\(^a\)City University, Department of Economics, D110 Social Sciences Bldg, Whiskin St., London EC1R 0JD  
\(^b\)City University, Department of Economics, D304 Social Sciences Bldg, Whiskin St., London EC1R 0JD

**Abstract**

We study the optimal execution strategy of selling a security. In a continuous time diffusion framework, a risk-averse trader faces the choice of selling the security promptly or placing a limit order and hence delaying the transaction in order to sell at a more favourable price. We introduce a random delay parameter, which defers limit order execution and characterizes market liquidity. The distribution of expected time-to-fill of limit orders conforms to the empirically observed exponential distribution of trading times, and its variance decreases with liquidity. We obtain a closed-form solution and demonstrate that the presence of the lag factor linearizes the impact of other market parameters on the optimal limit price. Finally, two more stylised facts are rationalised in our model: the equilibrium bid-ask spread decreases with liquidity, but increases with agents risk aversion.

**JEL classification:** D4, D81, D84, G1, G12

**Keywords:** order submission, execution delay, first passage time, risk aversion, liquidity traders

1. Introduction

The problem of optimal order placement is the kernel of the successful implementation of an investment strategy since the optimal trade execution reduces the associated transaction costs and augments expected returns. Traders construct their submission strategies to benefit from particular market properties and order types hence the architecture of the market defines their expectations about the future price dynamics and trade execution efficiency. In pure quote-driven dealer markets small orders typically execute at the best opposing dealer quote regardless of the order type. In public limit order books, market orders may encounter price improvement, whereas limit orders execution is conditional upon where traders place their limit prices relative to the prevailing bid and ask. The goals pursued by traders operating in a competitive environment of non-intermediated limit order markets are distinct from those of market makers. The fundamental distinction between the models of dealer quoting and that of a generic trader problem is that the former is essentially indifferent to execution, therefore his objective function is a zero profit. In limit order markets agents have multiple reasons to trade and various financial instruments available to execute a particular trade. Primarily, however, each trader entering a non-intermediated market must decide upon the type of order to use – a market order or a limit order.

Substantial empirical evidence suggests limit orders play a dominant role in the markets. For instance, Harris and Hashbrouck (1996) found that limit orders generally perform best even in the presence of a non-execution penalty and market price improvement. According to Guéant et al. (2011), more than half of all trades, approximately 60%, are as passive as possible, that is they fill the queue to trade rather than consume the liquidity. Blais et al. (1995) present an empirical analysis of the order flow of the Paris Bourse which is a pure limit-order market. They find that traders’ strategies vary with market conditions, with more limit orders at times when spreads are wide and more market orders

\(^*\)Corresponding author. Tel: + 44 (0)20 7040 4645, Fax: + 44 (0)20 7040 8580.  
Email addresses: polina.kovaleva.1@city.ac.uk (P. Kovaleva), g.iori@city.ac.uk (G. Iori)
at times when spreads are narrow. In real markets the motivation to choose a limit order over a market order is obfuscated by various subtle effects, for example: the discrepancy in transaction fees, the option to submit multiple orders simultaneously, the existence of several distinct markets for the same asset and the possibility to withdraw the order. Many previous studies were based on the division of the traders into two main groups: liquidity suppliers who trade with limit orders, and liquidity demanders who have higher immediacy priority. In contrast, we develop a framework where depending on the present market conditions the trader sets out either to provide or to consume liquidity.

In practice, even though investment and trading decisions are formulated simultaneously, they are usually analyzed and accomplished somewhat in isolation. Our model reflects precisely such allocation of tasks since we do not take into account the portfolio the trader might hold and how the outcome of the trading operations affects the balance and the value of this portfolio. Instead, we assume that the agent enters the market with a given trading goal and his task is to devise an optimal execution strategy given market characteristics at the time of his arrival. An important issue in market microstructure is the information set available to traders. There is abundant empirical evidence suggesting that market movements are often triggered by information updates and the presence of information asymmetries. A number of early studies showed that information component of the bid-ask spread is fractional; according to Huang and Stoll (1997), on average it compounds less than 12% of the spread. Yet, more contemporary findings indicate that the information component of the spread is significant and amounts for as much as 80% (Gould et al., 2010). Information asymmetry is usually defined in a fairly broad sense and it does not necessarily imply, for instance, any form of legal or illegal insider trading. Simply because the tools with which market participants assess the market vary, they draw different predictions from the same market conditions. However, superior or rather heterogeneous information is not the only basis for trading as was demonstrated by Milgrom and Stockey (1982). Guided by this line of thought, we do not analyze information effects per se; we incorporate traders’ individual expectations of the future asset price in our model, while leaving the grounds for this valuation beyond the scope of the present study.

In the present study we introduce a tractable parameter to characterize limit orders. More precise, we consider a problem of a trader who has to liquidate a position in an actively traded asset within a given period of time and forms his strategy based upon the market dynamics represented by bid and ask prices. Building upon the model in Iori et al. (2003), we conceptually improve it by incorporating an exogenous limit order execution factor – an exponential random delay. Moreover, we use a quadratic utility function and examine how different risk perception affects the optimal strategy of a trader in a limit order market and his average waiting time. This formulation proves more advantageous both in terms of interpretation and the ease of potential calibration to data. Further we provide an explicit static solution to the limit order trading problem for the quadratic utility preferences and subsequently identify the key determinants of the limit order attractiveness to the trader.

The remainder of this paper proceeds as follows. Section 2 reviews relevant theoretical and empirical literature. Section 3 describes the design of a market in which traders submit their orders and outlines the clearing mechanism. Section 4 presents the problem of a risk-averse trader who operates in this market and the expected payoffs associated with particular execution mechanisms. Section 5 reports comparative statics analysis for the parameters of the model. The subsequent Sections 6 and 7 analyse two special cases and the distribution of the waiting time respectively. The existence of equilibrium spread and the appropriate conditions are discussed in Section 8. Section 9 concludes.

2. Literature Review

There is an extensive literature on the subject of the optimal order submission strategy in limit order markets. The main distinction between the theoretical approaches adopted in various studies lies in the definition of the limit order execution mechanism and, consequently, the resulting probability distribution.

Equilibrium analysis of order-driven markets has been realized by Kumar and Seppi (1994); Chakravarty and Holden (1995); Parlour (1998); Foucault (1999); Foucault et al. (2005); Goettler et al. (2005) and Rosu (2009). All of these models are variants of a dynamic multi-agent sequential bargaining game where heterogeneous traders derive their best-response order submission strategies. Parlour (1998) assumes that the probability of execution of a sell
limit order depends on the arrival of buy market orders and the relative attractiveness of buy market orders depends on relative attractiveness of buy limit orders, thus execution probabilities of buy and sell limit orders are determined jointly over time. In situations when prices are fixed, and that should be the case in equilibrium, optimal order placement is contingent upon a single factor – the distribution of agents’ impatience characteristic. Allowing for price movements, Foucault (1999) describes the asset price via binomial model and assumes that limit orders are valid only for one period, therefore, at each point in time the book is either full or empty. In this setting the probability of limit order execution is endogenous and part of execution risk arises from the next trader’s order type. The focus on the optimal behavior in equilibrium yields numerous implications coherent with the documented market observations and these models proved especially useful for policy-makers. However, order-driven markets do not seem ever to attain these conditions: while in equilibrium all market participants should get zero profits, the depth of real limit order book is usually insufficient to drive average expected profits to zero.

A separate branch of optimal order placement literature, where the individual traders’ order submissions are aggregated by the asset price dynamics, was initiated by Cohen et al. (1981). In their seminal paper, Cohen et al. (1981) propel the theoretical analysis of optimal choice between market and limit orders in a framework with the probability of order execution contingent upon future price movements and the associated probability densities. Trading takes place when the trajectory of the best quote first crosses the barrier determined by the limit price. Cohen et al. (1981) model the security price with a compound Poisson process, which, by the very definition, invokes a jump in the probability of execution: if a time-constrained trader is willing to buy a stock via limit order and sets a price infinitely close to the current best ask, the probability of trading never attains unity. This property permitted to establish a so-called “gravitational pull effect”: when the bid-ask spread is narrow, the benefit of a price improvement with a limit order becomes small compared to the risk of non-execution so traders are pulled to use market orders instead. Consequently, Cohen et al. (1981) argue that limit order strategy is not always superior to trading with market orders or not submitting any orders at all. This model adequately captures the trade-off between favorable price and order execution probability and, in this sense, draws a line between market and limit orders. Cohen et al. (1981) further demonstrate that as the order arrival rate increases the Poisson process converges to the Wiener process eliminating the discontinuity in the execution probability function. The model setup, however, is too complex to obtain a closed-form solution and their analysis remains qualitative for the main part.

Langnau and Punchev (2011) concentrate deliberately on the issue of adequate price modeling in a public limit order book. Adopting the results achieved by Kou and Wang (2003), they compare the implications of a pure diffusion and a double exponential jump diffusion mid-point price generating processes. The most appealing properties of the double exponential jump distribution include the memorylessness of the price and the ability to accommodate the leptokurtic nature of returns. They arrive at the closed-form solution for the first passage time with distinct expressions for buy and sell strategies in the latter case. The compelling result of this paper is that the jump diffusion specification accommodates the asymmetric shape of a limit order book as well as a fat-tailed distribution of log returns. Therefore, Langnau and Punchev (2011) establish that exponential double jump diffusion specification provides a more accurate description of the markets and is compatible with the equilibrium conditions whereas log-normal prices are not. The jump diffusion setting requires to calibrate or fit considerably more parameters thus impeding its practical implementation.

Given the analytical complexity of the setting with the jumps in prices it seems reasonable to address the problem of optimal strategy within the context of a continuous time diffusion price. Several contributions can be found in the financial and econometric literature. For instance, Iori et al. (2003) show that despite its obvious shortcomings, the log-normal price still suggests that optimal limit order strategy is coherent with the traders behavior observed from the market. A mean-reverting price specification, particularly relevant for commodities, depicts an interesting cross-over effect: the optimal value of the strategy increases with the speed of reversion for small expiry times, while decreasing for big expiry times. Nevertheless, pure first passage time models such as the latter cannot justify the existence of spread due to the inability to differentiate between a marketable limit order and a market order, or more precisely, to distinguish among the time-to-first-fill, time-to-completion, and time-to-censoring when a limit order is withdrawn.

An attempt to preserve the appealing mathematical lightness of first passage time models and bridge it with the notion of imperfect liquidity was made by Harris (1998). Harris (1998) improves the framework with a pure diffusion

---

2 The orders placed inside the current spread are also interpreted as limit orders.
prices by adding a supplemental criterion – an aggregated factor of degree of execution difficulty. The degree of difficulty in limit order execution is defined as an additional barrier which a limit price has to pass before a limit order is filled. Therefore it is not sufficient to become the best price on the same side of the market, a limit order has to supersede this best price: for instance, a sell limit order is executed when submitted limit order price is lower than best ask less the difficulty parameter. Two execution mechanisms were juxtaposed: with certainty and with some probability. Comparative statics analysis based on a numerical solution confirms that the probability of order execution depends positively on this difficulty factor, in the certainty case as well. An alternative stylized interpretation of the problem was given by Hasbrouck (2006). Assuming that traders on the opposite side of the book are ascribed with unobserved reservation prices, a random collateral barrier following exponential distribution is analyzed. Hasbrouck (2006) argues that the number of potential counterparties is decreasing in the intensity rate of reservation prices distribution, thus optimal strategy becomes more aggressive. Although very valid, theoretical approaches presented by Harris (1998) and Hasbrouck (2006) have certain limitations. It appears to be a challenging task to infer the reservation prices or the degree of difficulty in limit order distribution from data due to the apparent difficulty in disentangling the impact of these external factors from the order aggressiveness determinants and the actual trading times.

Empirical research also reveals the bilateral causality in the relation between execution time and limit order submissions. Using survival analysis Lo et al. (2002) demonstrate that limit order execution times are increasing as the limit prices become more aggressive. This is both an outcome of the price priority rule and the potentially endogenous effect of aggressive order creating higher demand for trade from the other side of the book, as suggested by Parlour and Seppi (2008). Moreover, Lo et al. (2002) estimate the cumulative probability densities of the actual limit order execution times and compare to their hypothetical counterpart calculated as the first hitting times of geometric Brownian motion. The latter appears to understate largely expected trading times. Histograms of time-to-execution for limit orders exhibit exponential distribution and a comparison reveals that they differ not only in one or two moments but over their entire support. Tkatch and Kandel (2006) apply a simultaneous equations methodology to the Tel Aviv Stock Exchange data and find a significant causal impact of expected execution time on investors’ decisions of which orders are submitted. Without analysing directly how a particular trader’s order placement affects the market, our model captures the statistical properties of trading times and relates the expected time-to-fill to order aggressiveness.

3. The Model

3.1. The Market

We consider the problem of an investor who has a position in a traded asset which has to be liquidated within a pre-specified time horizon. One option the agent has is to use a market order and trade at the best available price at once. Alternatively, he can submit a limit order and hope to trade at a more favorable price. In a double auction market a transaction occurs when a market order hits the quote on the opposite side of the book. We assume that there is no information asymmetry and future price dynamics depend only upon public information.

Market orders trade at the best price currently available at the market and is filled instantaneously. The actual execution price of a market order is subject to the current market situation – this aspect of trading is referred to as execution price uncertainty. On the one hand, if the trade is negotiable then the agent can obtain a price improvement: in a situation when the spread is wide a trader may gain the speed in order execution by stepping inside the spread and so compromising on the price. On the other hand, market order traders bear the risk of trading at worse prices than they initially expect due to the rapidly changing market environment.

Limit orders are instructions to trade at the best price available but only if it is not worse than the limit price specified by a trader. The probability that a particular order will be filled depends on its limit price. Marketable limit orders, being the most aggressive, are easiest to fill. There exist two main risks associated with limit order trading: execution uncertainty when market moves away from the submitted limit price hence the agent never trades and ex post regret when for various reasons prices move towards and through the limit price.

Without loss of generality we further focus on the problem of a limit sell order, therefore optimizing the limit ask price $K_a$. The position in a security must be liquidated within a fixed period of time $T$. Its current price is determined by the best bid $b_0$, the highest buying price, and the best ask $a_0$, the lowest selling price. A pair of stochastic log-normal
processes $b_t$ and $a_t$ describe the trajectories of the bid and ask prices in the book respectively:

$$da_t = \mu_a a_t dt + \sigma_a a_t dW^a_t$$  \hspace{1cm} (1)

$$db_t = \mu_b b_t dt + \sigma_b b_t dW^b_t$$  \hspace{1cm} (2)

where $W^a_t$ and $W^b_t$ are $\rho$-correlated standard Brownian motions, $E[dW^a_t dW^b_t] = \rho dt$.

We analyze short-term decisions when market conditions do not change substantially, so the assumption deterministic price trends is viable. Rosu (2009) observed through a theoretical dynamic model, that there is a co-movement effect between bid and ask: when the bid lowers after a sell market order arrives, the ask also decreases but by a smaller amount, therefore widening the spread. Therefore, we should expect to observe a positive correlation between the best bid and ask price movements in real markets.

3.2. Order Execution Criteria

When a trader implements a limit order and submits his preferred ask price $K_a$ to the sell side of the market, he is “competing” with other potential sellers, or the best ask process $a_t$. Placing a limit order far from the current quotes implies more severe competition and increases the chances that the opportunity to trade will not arise before expiry time. Intuition suggests that when the market is illiquid and there are many gaps in the book, moving away from current prices is less risky. Also, the probability of execution of more aggressive limit orders increases, since normally the gaps get filled if they approach the best price. Once all the offers to sell at a lower price are picked up by arriving buyers, then his limit sell price becomes the best ask in the market. This is a general operation principle of double auction markets, however, the subsequent order matching and clearing is a less trivial process. Apart from competing in price, traders also contend in order size and time priority. Consequently, for modeling purposes, we need to delineate some mechanical order execution rule which will take into account such aspects of trading.

The market we study is sufficiently liquid, implying that the price dynamics can be described by a continuous process. Limit orders in the book are executed in first-in, first-out time priority. Once the limit order of an agent becomes the best price in the market, a random delay before trading ensues. The delay is usually small relative to the time horizon, but there is a small probability that the delay will be sufficiently long and hence distort the schedule of the agent’s trading operations. The transaction occurs with an unforeseen delay $\varepsilon$ which is independent of the asset prices and is modeled as an exponential random variable with constant intensity $\lambda$. The delay in trading occurs due to the fact that somebody else might have put an order at exactly the same price but earlier than the agent in question. Another argument for delay, albeit a minor one in the context of modern electronic markets, is an operational delay. Clearly, in most of the cases, once this market is trading close enough to the quote that an agent has previously submitted, it will not move away swiftly. More importantly, this market design implies that the existence of a spread is related to the costs of waiting: even if the trader’s horizon is long enough the delay might force him to use a market order and pay the spread despite initially high chances of trading at a more favourable limit price.

In their empirical study based on survival analysis Lo et al. (2002) observe exponentially distributed trading times. The absence of the peak near zero can by partially attributed to the discrete nature of the data, whereas our result applies to continuous time. Cho and Nelling (2000) argue that given market orders arrive in a non-homogeneous Poisson process the waiting times of a limit order follow a Weibull distribution. The estimations from a duration model (based on TORQ data from Harris and Hasbrouck (1996)) suggest that a histogram of empirical observations resembling an exponential probability density as well as Weibull. Notably, our theoretical result is compatible with a Weibull distribution specification for a shape parameter smaller or equal to unity. Cincotti et al. (2005) use tick data from 7 different US financial markets, they find that the distribution of trade waiting times is a mixture of exponential processes and verify this implication on an agent-based artificial market model.

4. Agent’s Problem

The trader is alloted the task of selling one unit of the asset and has to complete the trade by a certain deadline. We assume that in order to optimize the price he is willing to adopt a limit order strategy but is aware that there is a penalty for the non-execution event. Therefore, we introduce the possibility of converting to a market order at maturity if his
limit order was not filled. If his limit order does not reach the front of the queue in the limit order book within the horizon \( T \), the agent has to sell the asset at the best available bid.

In a perfectly liquid market even if the initial limit order at a certain price \( K_a > a_0 \) was not picked before \( T \), submitting a sell limit order at the best ask guarantees trading immediately. Thus, a time-zero discounted payoff of the agent’s strategy equals the liquidation value of the asset:

\[
V(K_a, a_0; T) = e^{-\delta T} K_a I_{[T \leq t]} + e^{-\delta T} a_T I_{[t > T]},
\]

where \( \tau = \inf\{t \geq 0: a_t = K_a\} \). Essentially, in the perfectly liquid market, \( a_t = b_t \) must hold at all times because the execution mechanisms is exactly the same at both prices.

In a situation with a random delay the agent cannot bear additional risk at maturity. Hence submitting a limit order is not a sure trade and he is obliged to use market order instead and sell the security at the highest price on the buy side of the market at time \( T \) – that is the best bid \( b_T \). Suppose a trader decides to place a limit order inside the spread, then his only concern is a random delay matters.\(^3\) However, placing an order at the best ask entails same expected time-to-fill but at a more beneficial price. Therefore, in this stylised framework no trader will be willing to sell security even one tick cheaper than the current best ask\(^4\) This strategy yields the following expression for the discounted terminal payoff:

\[
V_{\lambda}(K_a, a_0, b_0; \lambda, T) = e^{-\delta(T + \varepsilon)} K_a I_{[T + \varepsilon \leq t]} + e^{-\delta T} b_T I_{[t > T]},
\]

with \( \tau = \inf\{t \geq 0: a_t = K_a\} \) and \( \varepsilon \sim Exp(\lambda) \).

Numerous studies solve for the optimal strategy from the stand-point of a risk-neutral agent. However, if risks cannot be hedged away, the trader is concerned not only about the expected payoff but also about the range where the future payoff might lay, therefore he should take into account the variance in the future wealth. We assume that the trader is ascribed with a \( \varphi \) degree of risk-aversion and maximizes a mean-variance utility function:

\[
EU(K_a) \equiv \max_{K_e < a_0} \left[ E[V] - \varphi \cdot Var[V] \right].
\]

**Proposition 1.** Denote the power function of the profit from selling a unit of security as \( G_\lambda(\gamma) = (V(K_a, a_0, b_0; \lambda, T))^\gamma \). Given that market prices are positively correlated log-normal processes (1) and (2), the expected value of \( G \) for selling at a limit price \( K_a \geq a_0 \) in a market with a random delay in limit order execution is given by expression:\(^5\)

\[
EG_\lambda(\gamma) = \mathbb{E}[\left( \frac{K_a}{a_0} \right)^{\lambda + \frac{\gamma \delta}{\sqrt{T}}} N\left( \frac{-x - A_2 \sigma^2 T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_{a_T}}{a_0} \right)^{A_1 - \lambda} N\left( \frac{-x + A_3 \sigma^2 T}{\sigma_a \sqrt{T}} \right)]
\]

\[
= e^{-(\lambda + \frac{\gamma \delta}{\sqrt{T}})} \mathbb{E}\left[ \left( \frac{K_a}{a_0} \right)^{A_2} N\left( \frac{x - A_2 \sigma^2 T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_{a_T}}{a_0} \right)^{A_1 - A_2} N\left( \frac{-x - A_3 \sigma^2 T}{\sigma_a \sqrt{T}} \right) \right]
\]

\[
= b_0^\gamma e^{\gamma(\mu + (\gamma - 1)\sigma^2/2 - \delta)} \mathbb{E}\left[ \left( \frac{K_a}{a_0} \right)^{A_1} N\left( \frac{x - A_1 \sigma^2 T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_{a_T}}{a_0} \right)^{A_1} N\left( \frac{-x - A_3 \sigma^2 T}{\sigma_a \sqrt{T}} \right) \right]
\]

\[
= e^{-\gamma T} \left( \frac{K_a}{a_0} \right)^{A_1 + A_2} N\left( \frac{x - A_1 \sigma^2 T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_{a_T}}{a_0} \right)^{A_1 - A_2} N\left( \frac{-x - A_3 \sigma^2 T}{\sigma_a \sqrt{T}} \right)
\]

where \( x = \ln\left( \frac{K_a}{a_0} \right) \) and the constants are calculated as follows:

\[
A_1 = \frac{\mu - \sigma^2/2}{\sigma^2}, \quad A_2 = \sqrt{\left( \frac{\mu - \sigma^2/2}{\sigma^2} \right)^2 + 2\rho \sigma^2}, \quad A_3 = \sqrt{\left( \frac{\mu - \sigma^2/2}{\sigma^2} \right)^2 - 2\rho \sigma^2}.
\]

\[
A_4 = \frac{\mu - \sigma^2/2 + 2\rho \sigma \sigma^2}{\sigma^2}, \quad A_5 = \sqrt{\left( \frac{\mu - \sigma^2/2 + 2\rho \sigma \sigma^2}{\sigma^2} \right)^2 - 2\rho \sigma^2}.
\]

\(^3\)Mike and Farmer (2008) even argue that in determining a trade-off between the price improvement and execution risk, the distributions of limit prices in the book and inside the bid-ask spread differ qualitatively. The reason is that the first-in-first-out rule does not apply in the latter case.

\(^4\)We actually impose in this way that time increments are not equally spaced but the intervals between the price moves.

\(^5\)The mean and the variance of the expected profit from selling the asset are: \( EV_{\lambda} = EG_\lambda(1) \) and \( Var[V_{\lambda}] = EG_\lambda(2) - (EG_\lambda(1))^2 \).
with the parameter constraints \( \lambda \in \left[ 0; \frac{(\mu_a - \sigma_a^2/2)^2}{2\sigma_a^2} \right] \) and \( \rho \geq 0 \).

**Proof.** See Appendix A.1. \( \square \)

Consider the case when a limit order to sell is submitted exactly at the current ask. The expected value of this strategy is \( EV_{||K_a=a_0} = a_0 \left( 1 - e^{-(1+\delta)T} \right) + b_0 \left( 1 - e^{-(1+\delta)T} \right). \) The greater is the average delay \( E[\tau] = 1/\lambda \), the higher is the weight given to the profit from using a market order at maturity. If \( \lambda = 0 \), that is, the delay is substantially long, then the expected payoff is exactly the discounted expected bid \( EV_{||K_a=a_0} = e^{-\delta T} E [b_T] \).

Regarding the upper bound of delay parameter, an intuitive interpretation for this benchmark lies in its relation to the risk-adjusted mean squared to the variance of \( \tau \), which is referred to the squared Sharpe ratio. This upper bound is rather low for realistic values of the parameters and implies a strong liquidity restriction, however, qualitatively the results appear to be very similar to those based on formula (7).

**Proposition 2.** Denote the power-law function of the profit from selling a unit of security as \( G_a (\gamma) \equiv (V(K_a, a_0, b_0; \lambda, T))^\gamma \). Given the market prices are \( \rho \)-correlated log-normal processes (1) and (2), the expected value of function \( G \) for selling at a limit price \( K_a \geq a_0 \) in a market with a small random delay in limit order execution converges to the value:

\[
\lim_{\lambda \to \infty} EG_a (\gamma) = \left( \frac{K_a}{a_0} \right)^{A_1 + A_2} N \left( \frac{-x - A_2 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_a}{a_0} \right)^{A_1 - A_2} N \left( \frac{-x + A_2 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) + b_0 \gamma e^{(\mu_a + (y-1)\sigma_a^2/2-\delta)T} \left[ N \left( \frac{x - A_4 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) \right] \left( \frac{K_a}{a_0} \right)^{2A_1} N \left( \frac{-x - A_4 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right)
\]

where \( x = \ln \left( \frac{K_a}{a_0} \right) \) and \( A_1 = \frac{\mu_a - \sigma_a^2/2}{\sigma_a^2}, \ A_2 = \frac{\mu_a - \sigma_a^2/2 + \gamma \sigma_a \sigma_\gamma}{\sigma_a^2}, \ A_4 = \frac{\mu_a - \sigma_a^2/2 + \gamma \sigma_a \sigma_\gamma}{\sigma_\gamma^2} \).

**Proof.** See Appendix A.2. \( \square \)

Notably, this result replicates the valuation formula derived in Iori et al. (2003) for the market where limit orders trade upon achieving the beginning of the queue on the relevant side. However, in accordance with the payoff in equation (3), in an infinitely liquid market a trader will resolve to a limit order at \( a_T \) in the non-execution event rather than pick the best buying order at \( b_T \). Nonetheless, as we show later, this setting actually requires a zero bid-ask spread, or \( a_i = b_i \), \( \forall i \).

In order to assess approximately the expected utility we compute its upper and lower bounds.\(^6\) Fig. 1 captures the convergence of the actual expected utility value and the boundary conditions as distance-to-fill of a limit order increases. While for a risk-neutral traders the upper and lower bounds appear to be symmetric (Fig. 1a), this is not the case for positive \( \varphi \) where the lower bound becomes significantly nonlinear and more restrictive (Fig. 1b). The reason for the latter is that for more aggressive orders (when \( K_a/a_0 \) is small) the probability of limit order execution is high and expected time to reach the beginning of the queue is relatively short, thus the approximation of penalty term (B.6) yields a relatively small value and (B.8) – a relatively large.

5. **Comparative Statics**

5.1. **Baseline Parameters**

Using analytical expressions (7) and (6) the expected payoff in a perfectly and imperfectly liquid markets respectively, we find the optimal limit price which a trader should submit to achieve the maximum level of expected mean-variance utility. Further, we implement numerical integration with recursive adaptive Simpson quadrature rule to obtain the solution to the optimization problem for medium levels of liquidity (\( \lambda < \bar{\lambda} \ll \infty \)).\(^7\) We then examine the impact of various model parameters on the optimal decision of a trader.

---

\(^6\)The derivations are given in Appendix B.2.

\(^7\)\( \bar{\lambda} = \min \left\{ \frac{(\mu_a - \sigma_a^2/2)^2}{2\sigma_a^2}, \frac{(\mu_a - \sigma_a^2/2 + \gamma \sigma_a \sigma_\gamma)^2}{2\sigma_\gamma^2} \right\} \)

7
Figure 1: The expected utility of a sell limit order strategy for as a function of the limit price: (a) for a risk-neutral agent ($\varphi = 0$), (b) for a risk averse trader ($\varphi = 0.01$). Results are shown for $a_0 = 1000$, $b_0 = 995$, $\Delta = 1$, $T = 5$ days, $\mu = 10\%$, $\sigma = 20\%$, $\rho = 0.1$, $\lambda \rightarrow \infty$.

For the purposes of compatibility and comparison with the previous work we set the time-zero best ask in the book equal to $a_0 = 1000$ with the initial spread of 5. The tick size appropriate for this price range is $\Delta = \frac{1}{8}$, and we choose a time horizon $T = 5$ days in a benchmark case. The discount factor in this market is $\delta = 5\%$ per annum while the expected drift parameters are $\mu = \mu_a = \mu_b = 10\%$ and the annual volatility of $\sigma = \sigma_a = \sigma_b = 20\%$ with a small positive correlation in prices $\rho = 0.1$. It follows immediately from the assumption of no information asymmetry that all traders are equally informed about the value of the security. The bid and ask prices represent valuation on demand and supply sides of the same asset and should not diverge significantly at any point in time. Given these considerations, we work with the cases when the drifts of bid and ask are the same in order to preserve stationarity of the spread.\footnote{For instance, according to BATS tables the common tick size for FTSE 100 and FTSE 250 excluding high liquidity segment stocks trading in a price range between £1,000 and £4,999 is £1.}

The delays are characterized by the liquidity rate $\lambda = 8$ per day\footnote{Condition $\mu_a = \mu_b$ implies that a process $b_t/a_t$ is a martingale and $E\left[\frac{b_t}{a_t}\right] = \frac{b_0}{a_0}$.} and risk-aversion parameter is $\varphi = 0.01$ if not specified otherwise.

5.2. Comparative Static Effects

Prior to the discussion of the effects of various market parameters on the optimal order placement we look at the shape of the expected quadratic utility function. There are three possible market patterns: sideway market, up-trend market and down-trend market. A straightforward interpretation of the drift is to link it to the extent of traders’...
optimism or pessimism about the future price movements based on the past market performance. The greater the expected surge in the asset price, the higher the chances of a sell limit order getting filled before expiry. Our analysis concerns trading in a risky asset, so a drift higher than the discount factor $\delta = 5\%$ implies a rise in stock price. This, in turn, makes limit orders more attractive for the investor who is willing to sell.

![Figure 2](image_url)

Figure 2: The expected utility of a sell limit order strategy for a risk-neutral trader ($\varphi = 0$) as a function of the limit price: (a) no delay in execution, (b) in presence of a random delay in limit order execution. Results are shown for $a_0 = 1000$, $b_0 = 995$, $\Delta = 1$, $T = 5$ days, $\sigma = 20\%$, $\rho = 0.1$, $\lambda = 8$.

We compare the expected utilities of a risk-neutral (Fig. 2) and risk-averse (Fig. 3) traders. The expected profit that a risk-neutral agent attains trading in a perfectly liquid limit order book is the same for any limit price he might choose to submit (Fig. 2a) if the expected return of the underlying security equals risk-free rate. If the price trend is downward sloping – the best choice would be to sell at the current prevailing price, while upward trend implies that a trader should use an infinitely high price. This relationship matches perfectly the result obtained by Iori et al. (2003).

In contrast, the expected utility of a trader selling via limit orders in an illiquid market retains a concavo-convex shape for a range of risk aversion coefficients, including $\varphi = 0$ (Fig. 2b and Fig. 3b). Arguably, placing a limit order close to the current quote does not eliminate entirely a non-execution risk because of a random delay. On the one hand, moving away from current ask reduces the probability that the market will reach that level. On the other hand, given that correlation $\rho$ is small, the price of the market order at maturity is likely to be lower, therefore, the penalty is less tangible. Also, as Harris and Hasbrouck (1996) confirm, the opportunity cost of unfilled limit orders behind the best quotes is significant.

Focusing on the effect of the drift parameter we notice that the optimal $K^*_{a}$ in a situation without delay and no risk aversion is much more sensitive to the change in the price drift $\mu$. In effect, the conclusion to be drawn from Fig. 2a and Fig. 3a is that perfect liquidity implies binary choice: either to trade at the current quote or post an infinitely high limit sell price. The threshold $\hat{\mu}$ after which an agent switches from a marketable order to a more passive one is predictably moving rightward as risk aversion raises: it is below zero for $\varphi \leq 0.02$ (Fig. 4). Thereupon, in a market
with imperfect liquidity a trader tends to choose a more passive strategy for certain values of the price drift, while the optimal decision of the same trader in an absolutely liquid market under the same circumstances is not uniquely defined. Once a random delay is introduced to the market, the optimal limit price $K^*_a$ increases linearly with $\mu$, as Fig. 4 clearly reveals, while perfect liquidity assumes limit prices $K^*_a = \{a_0 \text{ or } \infty\}$. A possible explanation is that an investor operating in a liquid double auction will take on an opportunity with a limit order as soon as there appears a slight chance to make higher profit $\mu > \delta$ since the penalty which non-trade event implies is somewhat artificial – the worst case scenario is to submit a limit order at maturity at $a_T$ (Iori et al., 2003). The situation with delay offers a more compelling description of trading strategy (Fig. 4), suggesting a gradual increase in limit price with the increase in price trend.

The limit price is diminishing as the bid-ask spread at the time of order placement increases Fig. 5. Using Paris Bourse order flow data, Biais et al. (1995) find that limit orders prevail at times of wide spreads and market orders – at times of narrow spreads. This tendency is compatible with the trading pattern of arbitrageurs and high frequency traders. However, the validity of our result is in line with Ranaldo (2004) who observes that order aggressiveness of patient traders increases as spread widens.

Fig. 6 demonstrates the effect of delay intensity on $K^*_a$. The more risk-averse the trader, the smaller is an absolute impact of delay characteristic on the optimum. For relatively high levels of risk tolerance ($\varphi = \{0; 0.01; 0.02\}$) there is a positive relationship between liquidity $\lambda$ and the optimal limit price $K^*_a$. This is an intuitively appealing result: an increase in intensity rate $\lambda$ implies a decrease in mean delay (as well as its variance), thus, other things being equal, leads to a shorter time-to-completion of a trade. The results for the different levels of risk tolerance suggest that an increase in $\varphi$ decreases the optimal $K^*_a$, as expected.

Further, we examine the effect of the correlation between the paths of the best bid and ask prices. In a perfectly
liquid market, there is no bid-ask spread \((a_t = b_t)\) and the correlation is always unity. In a market with a random delay higher correlation coefficient reduces gradually the optimal limit sell price the agent should submit (Fig. 7). If ask price goes up then a limit order is more likely to get filled other things being equal. But if ask goes down and also bid goes down then will have to sell at a lower price. Therefore it makes sense that high \(\rho\) implies more passive order submissions. This suggests a certain borderline value of \(\rho\) when one effect outweighs the other. When the drifts are equal and there is a perfect correlation in bid and ask trajectories, we essentially deal with the situation of a constant spread. It follows that the plot for a constant bid-ask spread case corresponds to that of no spread in Iori et al. (2003) due to the fact that it becomes irrelevant.

Lastly, the order aggressiveness decreases as the maturity extends (Fig. 8). This result confirms previous statements that longer time horizons imply higher chances that \(a_t\) reaches the barrier represented by the trader’s limit price \(K_a^\ast\). Again, this proves the consistency of our theoretical framework. Moreover, the optimal strategy of a risk averse trader in an illiquid market is apparently less responsive to a change in expiry time.

6. Special Cases

We focus on two important special cases that allow us to separate two types of risks inherent to limit orders. First, we examine the situation when the time constraint is removed. Second, we look into the limiting behavior of the quotes chosen by the trader when the volatility of a traded asset is extremely low.

6.1. Infinite Time Horizon

Consider the case of infinitely long time horizon. We expect to detect a monotonic increasing relationship between the limit price \(K_a\) and the expected return. The intuition behind this reasoning is that once the time pressure is removed, the penalty which incurs paying the spread at maturity, simply vanishes. More precisely, we obtain

\[
\lim_{T \to \infty} EG_\lambda(\gamma) = \frac{A}{A + \gamma \delta} K_a a_0^{A_1 - A_2(\gamma)},
\]

(8)
Figure 5: The optimal limit sell price for a risk-averse trader as a function of initial spread $s_0$ (in ticks). Results are shown for $a_0 = 1000$, $b_0 = a_0 - s_0$, $\Delta = 1$, $T = 5$ days, $\mu = 10\%$, $\sigma = 20\%$, $\rho = 0.1$.

Figure 6: The optimal limit sell price for a risk-averse trader as a function of market liquidity (log$_2$-scale). Results are shown for $a_0 = 1000$, $b_0 = 995$, $\Delta = 1$, $T = 5$ days, $\mu = 10\%$, $\sigma = 20\%$, $\rho = 0.1$. 
Figure 7: The optimal limit sell price for a risk-averse trader as a function of the bid and ask correlation in presence of a random delay in limit order execution. Results are shown for $a_0 = 1000$, $b_0 = 995$, $\Delta = 1$, $T = 5$ days, $\mu = 10\%$, $\sigma = 20\%$, $\lambda = 8$.

Figure 8: The optimal limit sell price for a risk-averse trader as a function of the expiry time in presence of a random delay in limit order execution. Results are shown for $a_0 = 1000$, $b_0 = 995$, $\Delta = 1$, $\mu = 10\%$, $\sigma = 20\%$, $\rho = 0.1$, $\lambda = 8$. 

13
6.2. No Price Uncertainty

When the volatility of a security which the trader has to liquidate approaches zero, the price of this security will be changing at a constant rate per unit of time. Therefore, the maximum that the best ask can attain until maturity is known to be \( a_T = a_0 e^{\mu T} \), and the best bid is \( b_T = b_0 e^{\mu T} \).

\[
\lim_{\sigma \to 0} \mathbb{E} G_A(y) = \frac{\lambda}{\lambda + \gamma \delta} K_a^\gamma (e^{-\gamma \delta T} - e^{-\lambda \gamma \delta T}) + b_0 e^{(\gamma \mu - \gamma \delta - \lambda) T} \cdot e^{\lambda T}.
\]

(9)

The moment when the limit sell order \( K_a \) will hit the quote is calculated as \( \tau = \ln\left(\frac{K_a}{a_0}\right) \). In this situation the trader bears no price risk and the only risk he faces is linked to the non-execution of his orders due to delays. Considering the deterministic nature of prices, the trader will optimize his strategy only over the subset of limit prices from the interval \( a_0 \leq K_a \leq a_0 \cdot e^{(\mu - \delta) T} \) to ensure that \( \tau \leq T \).

7. The Distribution of Waiting Times

The probability density function of a time-to-fill, as depicted in Fig. 9b, follows an exponential distribution with a peak. The relationship between the distribution of a trading time in a market without delay and in a market with a random delay is stated in the following proposition.
Proposition 3. Let \( P(\tau \leq t) \) be the cumulative distribution of limit order time-to-fill in a perfectly liquid market, then the cumulative distribution function of waiting times in a market with small average delay \( P(\theta \leq t) \) is approximately

\[
P(\tau \leq t) - P(\theta \leq t) = \bar{\varepsilon} \frac{x}{\sigma_d \sqrt{t}} n \left( \frac{x - A_1 \sigma_d^2 t}{\sigma_d \sqrt{t}} \right).
\]

The relationship between the probability density functions for two types of markets is:

\[
P(\theta \in dt) = P(\tau \in dt) \left[1 - \bar{\varepsilon} \frac{x^2 - A_1^2 \sigma_d^4 t^2}{2 \sigma_d^2 t^2} - 3t\right].
\]

Proof. See Appendix A.3.

Since random delay \( \varepsilon \) follows exponential distribution, this proposition reveals that the discrepancy between waiting time in a market with infinite liquidity and a less liquid market is subject to an exponential component. The result complies with the intuition that the difference in limit order execution is most sensitive to market liquidity when medium-length maturities are concerned; both equations (10) and (11) show a negative dependence on a time factor.

In addition to that, as Figure 10a reveals, the average waiting time \( \bar{\theta} \) grows linearly with the distance-to-fill of a limit order. However, the contribution of the market liquidity parameter \( \lambda \) is negligible. Given that the distribution of the time-to-fill is considerably skewed, the mean is not an informative statistic. We plot the median of the time-to-fill \( \hat{\theta} \) – a-50%-probability outcome. For instance, if a trader operating in a perfectly liquid market submits a sell order at \( K_a = 1020 \), then it is equally likely to execute earlier or later than his horizon \( T = 5 \) days; whereas, in a market with
imperfect liquidity $\lambda = 1$ he must use a more aggressive order $K_a = 1017$ to achieve this. As Fig. 10b demonstrates, the median trading time increases as passive orders fill the book. Also, the higher is the liquidity (lower $\lambda$), the shorter is the median time-to-fill.

8. The Equilibrium Spread

It follows from the preceding discussion that in a continuous double auction the optimal decision of every market participant is made on the basis of existing best bid and ask prices which define the market spread. Meanwhile, the choices and actions that traders take now affect the current market spread and, as a direct consequence, influence future trading decisions. In fact, the market spread is the result of the interactions between heterogeneous agents who populate the market.

In the quote-driven markets the bid-ask spread is charged by the market maker in order to cover the expenses incurred by trading against better informed agents. The reasons for the bid-ask spread in a double-auction markets are more subtle; its presence is justified in several ways in the literature: information asymmetry which we have discussed before, gravitational pull effect depicted by Cohen et al. (1981), variations in the state of the book and traders valuations for the asset, trading costs, which in turn comprise direct costs such as commissions, transfer taxes order submission fees and account service fees, and indirect cost – the difference between the price at which the transaction was actually carried and a certain fair price. The delay parameter absorbs all these nuances which, in essence, make the difference between trading with a limit order rather than a market order. Cohen et al. (1981) examine this issue in detail and show that a positive market spread is incumbent to this market microstructure. According to Cohen et al. (1981), the equilibrium spread in a dynamic trading system is “the bid-ask spread at which, for the next instant of time, the probability of the spread increasing is equal to the probability of the spread decreasing.” Adopting his definition to our framework we arrive at the condition (see Appendix B.1.1):

$$\frac{\mu_a}{\mu_b} = \frac{b_t}{a_t}.$$ (12)

This condition, as Cohen et al. (1981) emphasize, does not guarantee that the market will eventually settle at this equality, rather that it is more likely that the price will move towards this condition than in the opposite direction. Since the drift parameters are constant, the condition in equation (12) simply states the equilibrium level of spread is constant and is determined by the gap between the growth rates of bid and ask prices. However, in order to preserve non-negative spread we must have $a_t \geq b_t$, $\forall t$, or $\mu_a \leq \mu_b$ in equilibrium. This slightly counter-intuitive contingency arises from the specific behavior of log-normal process – an increase and a decrease by the same amounts are not equally likely. If we impose for the purpose of stationarity $\mu_a = \mu_b$ then we must observe a zero absolute spread in equilibrium – equation (12) requires $a_t = b_t$, $\forall t$.

Nevertheless, this definition does not account for the notion of delay which is the crucial characteristic of the market we examine. It is precisely the deviation from the perfect liquidity case inflicted by large delays in limit order execution that determines the real spread size.

**Definition (Equilibrium Spread).** In a dynamic trading process the equilibrium market spread is the bid-ask spread such that the expected utility from trading via a limit order at the optimal limit price is equal to the utility from an immediate market order.\(^{11}\)

Applying numerical optimization we find the implied equilibrium spread given the expected delay.\(^{12}\) It has been discussed in the empirical literature that the main determinants of the spread size are competition for liquidity and risk aversion of market participants (Ranaldo, 2004). The higher is the competition among traders to provide liquidity, the tighter is the observed bid-ask spread, whereas the degree of risk aversion of the traders has a positive impact on the spread size. As depicted in Fig. 11a, the size of spread diminishes and eventually attains zero as the value of $\lambda$

---

\(^{11}\)This is an interpretation of a definition suggested by Harris (2003) (p.304): “The spread which ensures that traders are indifferent between using a limit order and a market order is the equilibrium spread.”

\(^{12}\)The necessary derivations are presented in Appendix B.1.2.
increases, thereby decreasing the expected delay. In Fig. 11b we can see that the equilibrium spread increases, as expected, with the degree of risk aversion $\phi$.

Although this result is obtained from the standpoint of a seller, the pattern for the buyer would be symmetrical in our setting. Moreover, since the limit orders are convertible in our model, seller has to monitor both sides of the book. In other words, a market design where a limit order is executed immediately the moment it becomes the best price in the market requires a zero bid-ask spread.

9. Conclusions

We have developed a logically consistent and empirically plausible model, which is easy to estimate. The central feature of the present study is the analysis of the impact of a random delay in limit order execution on the optimal strategy of a risk-averse trader. Based on an analytical solution, we examine the effects of various market parameters on his optimal selling strategy.

Our framework both benefits from transparency and explains the trade-off between immediacy and a favorable transaction price. In contrast with standard first passage time models of trading, it captures the fundamental difference between time required to reach the beginning of the queue on the relevant side of the market and the time-to-completion of a trade. The probability density of the expected time-to-fill of limit orders sharpens as liquidity increases and reveals an empirically observed exponential distribution of trading times. The discrepancy, as the model confirms, is due to imperfect liquidity, which in turn defers the trade execution. The main result suggests that the introduction of a random delay factor alleviates the impact of various market conditions on the optimal limit price the trader submits. Using comparative statics analysis, we demonstrate that the presence of a lag factor linearizes these effects. Notably, it is not the magnitude but the mere presence of delay that alters the nature of the relationship. Furthermore, we determine the equilibrium market spread as the bid-ask spread such that the expected utility from trading via a limit order at the optimal limit price is equal to the profit from an immediate market order. We subsequently prove that, consistent with real market phenomena, the equilibrium bid-ask spread increases both as liquidity decreases and agents’ risk aversion increases.

In addition to that we have demonstrated that in distinction from the profit-maximizing case, the introduction of risk-aversion factor provides the results which are more coherent with empirical observations and seem to be more useful for the practical implementation. The mean-variance utility function permits adequate risk assessment for a strategy involving limit order trading, therefore our approach allows to model the trading trajectories of heterogeneous investors.

We provide a solution for a static problem which can be extended to multi-period submission steps and solved in a manner of Harris (1998). However, we expect that results will not alter qualitatively once a trader is allowed to
revise his strategy a finite number of times. Whereas we analyzed only small trades, large trades should be examined differently since they have a price impact when market orders are used to execute the trade. There is a separate branch of literature on the order splitting issues which are closely related to our framework (Almgren and Chriss, 2000; Obizhaeva and Wang, 2005; Alfonsi et al., 2010; Løkka, 2011). In a recent paper by Guéant et al. (2011), the authors propose a novel approach of splitting a large trade using limit orders rather than market orders. This is the direction for the future development of our framework.

Appendix A. Proofs

Appendix A.1. Proof of Proposition 1

In order to find the analytical expression for the power of the limit order strategy payoff we need to calculate $EG_1(\gamma) \equiv E \left[ K_0^e e^{-\gamma t + \xi \tau} I_t \left( \tau + T \right) \right] + b_T^e e^{-\gamma T} I_{\tau + T \leq T_1}$. The cumulative distribution of the maximum of a Brownian motion with drift $\xi$ and volatility $\nu$ and $x \geq 0$ is equal to

$$P(M_x^\xi \leq x) = N\left( \frac{x - \xi T}{\nu \sqrt{T}} \right) - e^{\frac{\xi^2}{2}} N\left( \frac{x - \xi T}{\nu \sqrt{T}} \right), \quad (A.1)$$

and

$$P(\tau \in dt) = \frac{|x|}{\sqrt{2\pi \nu^2 T}} e^{-\frac{x^2}{2 \nu^2 T}}. \quad (A.2)$$

We want to calculate $J_1 \equiv E \left[ K_0^e e^{-\gamma (T + \tau)} I_{\tau \leq T} \right]$ and $J_2 \equiv E \left[ b_T^e e^{-\gamma T} I_{\tau + T \leq T_1} \right]$. Assuming that the sopping time and delays are independent and implementing the integrated expectations formula\(^{13}\) we arrive at the following expression:

$$J_1 = K_0^e E \left[ e^{-\gamma (T + \tau)} I_{\tau \leq T} \right] = K_0^e E \left[ e^{-\gamma T} I_{\tau \leq T} \right]$$

$$= K_0^e \int_0^T e^{-\gamma T} f(\tau) \int_0^{T-\tau} e^{-\gamma \tau} f(\tau) \, d\tau. \quad (A.3)$$

Since in our framework the prices follow geometric Brownian motion, the first passage time has an inverse Gaussian distribution with the probability density function $f(\tau)$ given in (A.2); the delay variable follows an exponential distribution with a positive parameter $\lambda$, then $f(\tau) = e^{-\lambda \tau}$. Therefore, we simplify the integral in (A.4):

$$\int_0^T e^{-\gamma T} f(\tau) \int_0^{T-\tau} e^{-\gamma \tau} f(\tau) \, d\tau = \int_0^T e^{-\gamma \tau} f(\tau) \left[ \int_0^\tau e^{-\gamma \tau} \lambda e^{-\lambda \tau} \, d\tau \right] d\tau$$

$$= \int_0^T e^{-\gamma \tau} f(\tau) \frac{\lambda}{\lambda + \gamma} \left[ 1 - e^{-\gamma (\lambda + \gamma) T} \right] d\tau$$

$$= \frac{\lambda}{\lambda + \gamma} \left[ \int_0^T e^{-\gamma \tau} f(\tau) \, d\tau - e^{-\gamma (\lambda + \gamma) T} \int_0^T e^{\gamma \tau} f(\tau) \, d\tau \right].$$

\(^{13}E(X) = E(E(X|Y))\)
Using equation (A.2) with $\nu = \sigma^2_a$ and $\xi = \mu_a - \sigma^2_a/2$ we rewrite the first term as

$$J_1 = K^2 \frac{\lambda}{\lambda + \gamma \delta} \left[ \int_{0}^{\tau} e^{-\gamma \delta \tau} \left( \frac{x - (\mu_a - \sigma^2_a/2)\tau}{\sigma_a \sqrt{\tau}} \right) d\tau - e^{-(A + \gamma \delta)\tau} \int_{0}^{\tau} e^{\gamma \delta \tau} \left( \frac{x - (\mu_a - \sigma^2_a/2)\tau}{\sigma_a \sqrt{\tau}} \right) d\tau \right].$$

Since the following equality holds for the normal density

$$e^{-\phi \lambda \nu \tau} = e^{-\phi \beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right),$$

with $\alpha = x/\sigma_a$, $\beta = -(\mu_a - \sigma^2_a/2)/\sigma_a$ and $\phi = \gamma \delta$ we rearrange the first component of $J_1$ as

$$\int_{0}^{\tau} e^{-\phi \lambda \nu \tau} \left( \frac{x - (\mu_a - \sigma^2_a/2)\tau}{\sigma_a \sqrt{\tau}} \right) d\tau = \int_{0}^{\tau} \frac{1}{\tau} \left( e^{\phi \beta \lambda \nu \tau} \left( \frac{x + A_2 \sigma^2_a \tau}{\sigma_a \sqrt{\tau}} \right) \right) d\tau,$$

where $A_1 = \frac{\mu_a - \sigma^2_a/2}{\sigma_a}$ and $A_2 = \frac{\sqrt{\mu_a - \sigma^2_a/2 + 2\gamma \delta \lambda \nu}}{\sigma_a}$.

Finally, using the identity

$$\int_{0}^{\tau} \frac{1}{\tau} \left( \frac{x + A_2 \sigma^2_a \tau}{\sigma_a \sqrt{\tau}} \right) d\tau = \frac{1}{\alpha} \left[ N \left( \frac{-\left|\alpha\right|}{\sqrt{\tau}} - \text{sgn}(\alpha) \beta \sqrt{\tau} \right) + e^{-2\phi \beta \lambda \nu} N \left( \frac{-\left|\alpha\right|}{\sqrt{\tau}} + \text{sgn}(\alpha) \beta \sqrt{\tau} \right) \right],$$

substituting $\alpha = x/\sigma_a$, $\beta = A_2 \sigma_a$ and $K_0 \geq a_0$, we get

$$\int_{0}^{\tau} \frac{1}{\tau} \left( e^{(A_1 + A_2)} \left( \frac{x + A_2 \sigma^2_a \tau}{\sigma_a \sqrt{\tau}} \right) \right) d\tau = \int_{0}^{\tau} \frac{1}{\alpha} \left[ N \left( \frac{-\left|\alpha\right|}{\sqrt{\tau}} - \text{sgn}(\alpha) \beta \sqrt{\tau} \right) + e^{-2\phi \beta \lambda \nu} N \left( \frac{-\left|\alpha\right|}{\sqrt{\tau}} + \text{sgn}(\alpha) \beta \sqrt{\tau} \right) \right].$$

The second integral in $J_1$ is modified via the following identity which holds for the values $\lambda \leq \beta^2/2$: \(^{14}\)

$$e^{\phi \lambda \nu \tau} = e^{-\phi \beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right).$$

Precisely, we obtain

$$\int_{0}^{\tau} e^{\beta \lambda \nu \tau} \left( \frac{x - (\mu_a - \sigma^2_a/2)\tau}{\sigma_a \sqrt{\tau}} \right) d\tau = \int_{0}^{\tau} \frac{1}{\tau} \left( e^{(A_1 + A_3)} \left( \frac{x + A_3 \sigma^2_a \tau}{\sigma_a \sqrt{\tau}} \right) \right) d\tau,$$

where $A_3 = \frac{\sqrt{\mu_a - \sigma^2_a/2 + 2\gamma \delta \lambda \nu}}{\sigma_a}$. Applying formula (A.6) with $\alpha = x/\sigma_a$, $\beta = A_3 \sigma_a$ and $K_0 \geq a_0$

$$\int_{0}^{\tau} \frac{1}{\tau} \left( e^{(A_1 + A_3)} \left( \frac{x + A_3 \sigma^2_a \tau}{\sigma_a \sqrt{\tau}} \right) \right) d\tau = \int_{0}^{\tau} \frac{1}{\alpha} \left[ N \left( \frac{-\left|\alpha\right|}{\sqrt{\tau}} - \text{sgn}(\alpha) \beta \sqrt{\tau} \right) + e^{-2\phi \beta \lambda \nu} N \left( \frac{-\left|\alpha\right|}{\sqrt{\tau}} + \text{sgn}(\alpha) \beta \sqrt{\tau} \right) \right].$$

\(^{14}\) $e^{\phi \lambda \nu \tau} = e^{\beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right) = e^{\phi \beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right) e^{-\phi \beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right) = e^{-\phi \beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right) = e^{-\phi \beta \lambda \nu \tau} \left( \frac{\alpha + \beta \nu}{\sqrt{\nu}} \right)$.
Combining the results (A.7) and (A.9) we arrive at the final expression for the first term:

\[
J_1 = K_{ua} \frac{A_{\text{bid}}}{A + \gamma \delta} \left[ N\left( \frac{-x - A_2 \sigma_2^2 T}{\sigma_\alpha \sqrt{T}} \right) + \frac{K_{ua}}{a_0} \right]^{A_{\text{ask}} - A_{\text{bid}}} N\left( \frac{-x - A_2 \sigma_2^2 T}{\sigma_\alpha \sqrt{T}} \right) + \frac{K_{ua}}{a_0} \right]^{A_{\text{bid}} - A_{\text{ask}}} N\left( \frac{-x - A_2 \sigma_2^2 T}{\sigma_\alpha \sqrt{T}} \right)
\]

Before calculating the second term we rewrite the bid and ask equations to express these processes in terms of a two-dimensional Brownian motion \((W_1, W_2)\):

\[
da_t = a_t \left( \mu_a dt + \sigma_a d\tilde{W}_t \right)
\]
\[
db_t = b_t \left( \mu_b dt + \sigma_b d\tilde{W}_t \right)
\]

where \(\sigma_a = (\sigma_a, 0)\) and \(\sigma_b = \left( \sigma_b, \sqrt{1 - \rho^2} \right)\). Therefore,

\[
a_t = a_0 e^{(\mu_a - \frac{1}{2} \sigma_a^2) t} \sigma_a T \tilde{W}_t,
\]
\[
b_t = b_0 e^{(\mu_b - \frac{1}{2} \sigma_b^2) t} \sigma_b T \tilde{W}_t.
\]

Notice that if \(b_t\) is a log-normal process then \(b_t^\gamma\) is also log-normally distributed. Applying Ito’s formula to this process we can show that

\[
db_t^\gamma = b_t^\gamma \left( \gamma \sigma_b + \frac{\gamma (y - 1)}{2} \sigma_b^2 \right) dt + \gamma \sigma_b \sigma_a d\tilde{W}_t.
\]

Hence, applying Ito’s lemma once again we get \(b_t^\gamma = b_0^\gamma e^{(\gamma \mu_a (y - 1) \sigma_a^2/2 - \gamma \sigma_b) T} \sigma_b T \tilde{W}_t\). Now we rewrite the second term

\[
J_2 = e^{-\gamma T} \mathbb{E} \left[ b_0^\gamma e^{(\gamma \mu_a (y - 1) \sigma_a^2/2 - \gamma \sigma_b) T} \sigma_b T \tilde{W}_t \cdot I_{\{T > T^-\}} \cdot (I_{\{T > T^-\}} + I_{\{T > T^+\}}) \right]
\]

where expectation is calculated under probability measure \(P^\star\) defined by Radon-Nikodym derivative \(\eta_T = \frac{dp}{dp^\star} = e^{\gamma \tilde{W}_T - \gamma \tilde{W}_T^\gamma T}\). As Girsanov theorem states, \(\tilde{W}_T^\gamma = \tilde{W}_T - \gamma \tilde{W}_T\) is a two-dimensional Brownian motion under \(P^\star\), so we rewrite

\[
X_T = (\mu_a - \sigma_a^2 / 2) T + \sigma_a \left( \tilde{W}_T - \gamma \tilde{W}_T \right) = (\mu_a + (y - 1/2) \sigma_a^2) T + \sigma_a \tilde{W}_T^\gamma,
\]

implying a drift \(\xi^\gamma = \mu_a - \sigma_a^2 / 2 + \gamma \rho \sigma_a \sigma_b\) under \(P^\star\).

The first integral in (A.11) is the probability that the first passage time exceeds horizon \(T\), which is given in formula (A.1) for \(\nu = \sigma_a^2\) and \(\xi^\gamma\):

\[
\int_T^\infty |x| \left( \frac{x - (\mu_a - \sigma_a^2 / 2 + \gamma \rho \sigma_a \sigma_b)}{\sigma_a \sqrt{T}} \right) d\tau = N\left( \frac{x - A_4 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) - \left( \frac{K_{ua}}{a_0} \right) \left( \frac{A_4 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right),
\]

where \(A_4 = \frac{\mu_a - \sigma_a^2 / 2 + \gamma \rho \sigma_a \sigma_b}{\sigma_a^2}\). Furthermore, simplifying the last term in (A.11) yields

\[
\int_0^T f(t) \int_T^\infty f(\varepsilon) d\varepsilon d\tau = \int_0^T f(t) \int_T^\infty \lambda e^{-\lambda t} d\varepsilon d\tau = \int_0^T f(t) e^{-\lambda(t - T)} d\tau = e^{-\lambda T} \int_0^\infty e^{\lambda t} f(t) d\tau.
\]
We substitute the probability density function $f(\tau)$ with volatility $\sigma_a$ and drift $\xi = \mu_a - \sigma^2_a/2 + \gamma \rho \sigma_s \sigma_b$ and using property (A.8) obtain
\[
\int_0^T e^{\tau T} \frac{1}{\sigma_a \sqrt{T}} \left( x - \left( \frac{\mu_a - \sigma^2_a/2 + \gamma \rho \sigma_s \sigma_b}{\sigma_a \sqrt{T}} \right) \right) d\tau = \frac{1}{\sigma_a \sqrt{T}} \int_0^T \left[ e^{(A_1 + A_2) \tau} \left( x + A_3 \sigma^2_a T \right) \right] d\tau, \quad \text{(A.13)}
\]
where $A_5 = \sqrt{\mu_a - \sigma^2_a/2 + \gamma \rho \sigma_s \sigma_b} \frac{\sqrt{2 \pi}}{\sigma^2_a}$ and $A_5 \leq \frac{\mu_a - \sigma^2_a/2 + \gamma \rho \sigma_s \sigma_b}{\sigma^2_a}$. Given the parameters values $\alpha = x/\sigma_a$, $\beta = A_5 \sigma_a$ and $K_0 \geq a_0$ we write an explicit expression for this integral as shown in formula (A.6)
\[
\frac{|x|}{\sigma_a} \int_0^T \frac{1}{\tau \sqrt{T}} \left[ e^{(A_1 + A_2) \tau} \left( \frac{x + A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right] d\tau = \frac{K_0}{a_0}^{A_1 + A_2} N \left( \frac{-x - A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_0}{a_0}^{A_1 - A_2} N \left( \frac{-x + A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right). \quad \text{(A.14)}
\]

Using (A.12) and (A.14),
\[
J_2 = \begin{cases} 
\frac{b_0 e^{(\mu_a + (\gamma - m) \sigma_a^2/2 - \delta) T}}{\gamma \rho \sigma_s \sigma_b T} \left[ \left( \frac{K_0}{a_0} \left( \frac{x - A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right) - \left( \frac{K_0}{a_0} \left( \frac{-x - A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right) \right] \\
+ e^{-\frac{\lambda}{\sigma_a} \gamma \rho \sigma_s \sigma_b T} \left[ \left( \frac{K_0}{a_0} \left( \frac{x - A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right) + \left( \frac{K_0}{a_0} \left( \frac{-x + A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right) \right] \end{cases}
\]

Thus,
\[
\text{E}G_{\lambda}(\gamma) = K_0 \frac{\lambda}{\beta} + \frac{\gamma \rho \sigma_s \sigma_b}{\sigma_a} \left[ \left( \frac{K_0}{a_0} \left( \frac{x - A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right) + \left( \frac{K_0}{a_0} \left( \frac{-x + A_3 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right) \right]
\]
and $\lambda = \min \left\{ \frac{(\mu_a - \sigma^2_a/2)^2}{2 \sigma^2_a}, \frac{(\mu_a - \sigma^2_a/2 + 2 \rho \sigma_s \sigma_b)^2}{2 \sigma^2_a} \right\}$.

**Appendix A.2. Proof of Proposition 2.**

We find the limit of the expected utility using Laplace’s integral approximation method. Define an integral
\[
I(\lambda) = \int_a^b h(\tau) e^{\lambda g(\tau)} d\tau, \quad \text{(A.15)}
\]
where $[a, b]$ is a finite interval and functions $h(\tau)$ and $g(\tau)$ are continuous.

**Lemma 1.** Suppose the function $g(\tau)$ attains a maximum on $[a, b]$ at either endpoint, $\tau_0 = a$ or $\tau_0 = b$, and is differentiable in a neighborhood of $\tau_0$, with $g'(\tau_0) \neq 0$ and $h(\tau_0) \neq 0$. Then the leading term of the asymptotic expansion of the integral (A.15), as $\lambda \to +\infty$, is given by
\[
I(\lambda) = \frac{h(\tau_0) \cdot e^{\lambda g(\tau_0)}}{\lambda g'(\tau_0)} \quad \text{(A.16)}
\]
We apply this lemma to functions $g(\tau) = \tau$ and $h(\tau) = \frac{\ln(x - A_2\sigma^2 T)}{\sigma^2}$ with integration limits $a = 0$, $b = T$. The function $g(\tau)$ attains the highest value at $\tau_0 = b = T$ and $g'(\tau_0) = 1$, therefore we obtain from formula (A.16):

$$\int_0^T e^{\lambda \tau} \left[\frac{|a|}{\sqrt{T}} e^{\alpha + \beta \tau} \right] d\tau = \frac{|a|}{\sqrt{T}} n \left(\frac{\alpha + \beta T}{\sqrt{T}}\right) + e^{\lambda \tau} \left[\frac{|a|}{\lambda T \sqrt{T}} \cdot n \left(\frac{\alpha + \beta T}{\sqrt{T}}\right)\right].$$

We use this result to approximate the second integral in (A.13) when $\lambda \geq \beta^2 / 2$ and identity (A.8) does not hold. Let $\bar{\epsilon} = E[\epsilon] = 1/\lambda$, then it is easy to show that for high values of $\lambda$ the power utility is approximately

$$EG_\lambda(\gamma) \approx K_\lambda \left[\frac{K_\lambda}{{a_0}} \right]^{A_1 + A_2} N \left(-\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right) + \left(\frac{K_\lambda}{{a_0}}\right)^{A_1 - A_2} N \left(-\frac{x + A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right)$$

$$+ \bar{\epsilon} \cdot e^{-\gamma T} \left(\frac{x}{\sigma^2 \sqrt{T}}\right) \cdot n \left(\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right) - \frac{b}{a} \cdot e^{(\mu + (\gamma - 1)\sigma^2 / 2 - \beta t)} \cdot n \left(\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right) + \bar{\epsilon} \cdot e^{-\gamma T} \left(\frac{x}{\sigma^2 \sqrt{T}}\right) \cdot n \left(\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right),$$

where $x = \ln\left(\frac{K_\lambda}{{a_0}}\right)$, $A_1 = \frac{\mu - \sigma^2 / 2}{\sigma^2}$, $A_2 = \frac{\sqrt{\mu^2 - \sigma^4 / 2} + \lambda \beta \sigma}{\sigma^2}$, $A_4 = \frac{\mu - \sigma^2 (2 + \gamma) / 2}{\sigma^2}$. Further, noting that constants $A_1$, $A_2$ and $A_4$ are independent of $\bar{\epsilon}$, in a perfectly liquid market we obtain

$$\lim_{\bar{\epsilon} \to 0^+} EG_\lambda(\gamma) = K_\lambda \left[\frac{K_\lambda}{{a_0}} \right]^{A_1 + A_2} N \left(-\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right) + \left(\frac{K_\lambda}{{a_0}}\right)^{A_1 - A_2} N \left(-\frac{x + A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right)$$

$$+ b_0 \cdot e^{(\mu + (\gamma - 1)\sigma^2 / 2 - \beta t)} \cdot n \left(\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right) - \frac{b}{a} \cdot e^{(\mu + (\gamma - 1)\sigma^2 / 2 - \beta t)} \cdot n \left(\frac{x - A_2\sigma^2 T}{\sigma^2 \sqrt{T}}\right).$$

**Appendix A.3. Proof of Proposition 3**

In the presence of a random delay in limit order execution we denote time-to-fill as $\theta = \tau + \epsilon$ and obtain its distribution:

$$P(\theta \leq t) = \int_0^t f(\tau) \int_0^{\tau - T} f(s) ds d\tau$$

$$= \int_0^t f(\tau) \left(1 - e^{-\lambda (\tau - T)}\right) d\tau$$

$$= \int_0^t f(\tau) d\tau - e^{-\lambda t} \int_0^t e^{\lambda \tau} f(\tau) d\tau$$

$$= P(\tau \leq t) - \bar{\epsilon} \cdot \frac{x}{\sigma^2 \sqrt{n}} \cdot n \left(\frac{x - A_1\sigma^2 T}{\sigma^2 \sqrt{n}}\right),$$

under on the probability space $(\Omega, E, F)$ with $x = \ln\left(\frac{K_\lambda}{{a_0}}\right)$, $A_1 = \frac{\mu - \sigma^2 / 2}{\sigma^2}$ and $\bar{\epsilon} = \frac{1}{\lambda}$.

Note that the probability density of first passage time $P(\tau \leq t)$ is defined in equation (A.2). Taking the derivative of the second term in expression (A.19) with respect to $t$, we arrive at the probability density function of limit order...
time-to-fill \( \theta \)

\[
P(\theta \in dt) = \frac{x}{\sigma_a \sqrt{\pi t}} e^{\left(\frac{x - A_1 \sigma_t^2 t}{\sigma_a \sqrt{t}}\right)} \int \left[1 - \frac{x^2 - A_1^2 \sigma_t^4 t^2 - 3t}{2 \sigma_a^2 t^2}\right] = P(\tau \in dt) \left[1 - \frac{x^2 - A_1^2 \sigma_t^4 t^2 - 3t}{2 \sigma_a^2 t^2}\right].
\] (A.20)

It follows immediately from (A.20) that when the expected delay approaches zero time-to-fill is equivalent to the time it takes to reach the front of the queue: \( \lim_{x \to 0^+} P(\theta \in dt) = P(\tau \in dt). \)

Taking into account that prices are independent of delays, we have \( E[\theta] = E[\tau + \epsilon] = E[\tau] + E[\epsilon]. \)

The mean of an inverse Gaussian given in (A.2) is \( x/\xi \), the mean value of exponential delays is \( 1/\lambda \), therefore we find the average waiting time-to-fill of a limit order:

\[
E[\theta] = \frac{\ln\left(\frac{\xi}{m}\right)}{\mu_a - \sigma_a^2/2} + \frac{1}{\lambda}.
\] (A.21)

The mode of the sum of two independent variable is a sum of their modes and since the mode of exponentially distributed random variable is zero, the mode of the time-to-fill is:

\[
Mode[\theta] = \frac{4 \ln\left(\frac{\xi}{m}\right)^2 (\mu_a - \sigma_a^2/2)^2 + 9\sigma_a^4}{2(\mu_a - \sigma_a^2/2)^2}.
\] (A.22)

Appendix B. Additional Derivations

Appendix B.1. The Equilibrium Spread Derivations

Appendix B.1.1. Derivation of Condition (12).

Denote \( S_t = a_t - b_t \), then the probability that the spread will increase in an infinitely small time increment \( dt \) is equivalent to saying that the change in ask will be greater than the change in bid, independently of the direction.

\[
Pr[da_t \geq db_t] = Pr\left[a_t(\mu_a dt + \sigma_a dW^a_t) \geq b_t(\mu_b dt + \sigma_b dW^b_t)\right]
\]

\[
= Pr\left[\frac{b_t}{a_t} (\mu_a dt + \rho \sigma_b \psi_1 + \sigma_b \sqrt{1 - \rho^2 \psi_2} \sqrt{dt}) \leq \mu_a dt + \sigma_a \psi_1 \sqrt{dt}\right]
\]

\[
= Pr\left[\frac{b_t}{a_t} (\mu_a - \sigma_a \psi_1) + \frac{b_t}{a_t} \sigma_b \sqrt{1 - \rho^2 \psi_2} \leq \left(\mu_a - \frac{b_t}{a_t} \psi_1\right) \sqrt{dt}\right]
\]

\[
= Pr[\psi_1 \leq z_1] = N(z_1),
\]

where \( \psi_1 \sim N(0, 1) \) are independent random variables, \( z_1 \equiv \frac{\mu_a - \mu_b + \rho \sigma_b}{\sqrt{\rho^2 \sigma_a^2 + \sigma_b^2}} \sqrt{dt} \) with \( \chi^2 = \frac{b_t}{a_t} \sigma_b^2 + \sigma_a^2 - 2 \sigma_a \sigma_b \frac{b_t}{a_t} \). It is easy to see that variance \( \chi^2 \) is strictly positive for non-zero volatilities of the asset prices \( \sigma_a \) and \( \sigma_b \) and \( |\rho| \leq 1 \).

Similarly, the probability of a narrower spread is expressed as

\[
Pr[da_t \leq db_t] = Pr[\psi_3 \leq z_3] = Pr[\psi_3 \leq -z_3] = N(-z_3) = 1 - N(z_3).
\]

Equating two expressions yields:

\[
Pr[da_t \geq db_t] = Pr[da_t \leq db_t] \quad \frac{\mu_a}{\mu_b} = \frac{b_t}{a_t},
\]

Appendix B.1.2. Numerical Calculation of the Spread in a Market with Delay.

The preferences of a trader are described by a mean-variance utility function (5). The equilibrium spread \( S \) is defined as a situation when the expected utility from using a limit order at the optimal price \( K^*_\gamma \) equals the utility of
the profit from trading via immediate market order. Let $\tilde{S} = a_t - b_t, \forall$, then
\[
\tilde{S} = a_0 - EU(K^*_a; \tilde{S}). \tag{B.1}
\]

Using the approximation from the previous section (A.19) we substitute $b_0 = a_0 - S$ and find an optimal limit price as a function of spread $K^*_a = K^*_a(S)$ for a range $S \geq 0$. Then we determine the pair of limit price and spread that satisfy condition (B.1).

Appendix B.2. Bounds of the Expected Payoff Power Utility

The general form of the integral we need to bound is
\[
\int_0^T e^{-\lambda(T-\tau)} f(\tau) d\tau, \tag{B.2}
\]
where the density is $f(\tau) = \frac{a}{\tau \sqrt{\pi}} n \left( \frac{\alpha + \beta \tau}{\sqrt{\tau}} \right)$ and $\lambda$ takes large values. It is easy to see that
\[
\forall \tau \in [0, T]: \quad e^{-\lambda(T+\tau)} \leq e^{-\lambda(T-\tau)} \leq 1,
\]
and since $f(\tau)$ is a bounded function we have
\[
\int_0^T e^{-\lambda(T+\tau)} f(\tau) d\tau \leq \int_0^T e^{-\lambda(T-\tau)} f(\tau) d\tau \leq \int_0^T f(\tau) d\tau.
\]

For the upper bound we apply identity (A.6)
\[
\int_0^T f(\tau) d\tau = \int_0^T \frac{|a|}{\tau \sqrt{\pi}} n \left( \frac{\alpha + \beta \tau}{\sqrt{\tau}} \right) d\tau
= N\left( \frac{-\alpha + \beta T}{\sqrt{T}} \right) + e^{-2|a|} N\left( \frac{-\alpha - \beta T}{\sqrt{T}} \right). \tag{B.3}
\]

In order to calculate the lower bound we first use property (A.5)
\[
e^{-\lambda(T+\tau)} f(\tau) = e^{-\lambda T} \frac{|a|}{\tau \sqrt{\pi}} \left[ e^{-\lambda T} n \left( \frac{\alpha + \beta \tau}{\sqrt{\tau}} \right) \right] = e^{-\lambda T} \frac{|a|}{\tau \sqrt{\pi}} \left[ e^{-\alpha T + \beta \tau + \frac{\beta^2}{2}} n \left( \frac{\alpha + \beta \tau}{\sqrt{\tau}} \right) \right],
\]
then property (A.6) to get
\[
\int_0^T e^{-\lambda(T+\tau)} f(\tau) d\tau = |a| e^{(\psi - \beta) - \lambda T} \int_0^T \frac{1}{\tau \sqrt{\pi}} n \left( \frac{\alpha + \psi T}{\sqrt{T}} \right) d\tau
= e^{-\lambda T} \left[ e^{(\psi - \beta) N} \left( -\alpha + \psi T \right) + e^{-\alpha T} \left( -\alpha - \psi T \right) \right], \tag{B.4}
\]
where $\psi = \sqrt{\beta^2 + 2 \lambda}$. 

24
The full expression for expected power utility is

\[ E_{\gamma} \alpha = \frac{K^y}{A + \gamma} \left[ \int_{0}^{T} e^{-\gamma^2 t} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 \right) t}{\sigma_a \sqrt{T}} \right) dt \right] \]

\[ \quad - e^{-(\lambda + \gamma) \delta T} \int_{0}^{T} e^{-\gamma^2 t} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 \right) t}{\sigma_a \sqrt{T}} \right) dt \]

\[ + \ b_0^y e^{(\mu_a + (\gamma - 1)\sigma^2_a / 2 - \delta) T} \left[ \int_{T}^{\infty} \frac{|x|}{\sigma_a^2} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 + \gamma \rho \sigma_a \sigma_b \right) t}{\sigma_a \sqrt{T}} \right) dt \right] \]

\[ + e^{-\gamma^2 T} \int_{0}^{T} e^{-\gamma^2 t} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 + \gamma \rho \sigma_a \sigma_b \right) t}{\sigma_a \sqrt{T}} \right) dt \].

As shown in the proof of Proposition 1, the first and the third integrals have analytical expression for all values of parameters. Since the second term in this expression is negative with \( \alpha = x / \sigma_a \) and \( \beta = -A_1 \sigma_a \) we apply formula (B.4) to determine its upper bound

\[ I_2 \leq -e^{-(\lambda + \gamma) \delta T} \left[ \frac{K_u}{a_0} \right]^{A_1 + A_6} N \left( \frac{-x - A_6 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_u}{a_0} \right)^{A_1 - A_6} N \left( \frac{-x - A_6 \sigma^2_a T}{\sigma_a \sqrt{T}} \right), \quad (B.5) \]

where \( A_6 = \sqrt{\frac{(\mu_a - \sigma^2_a / 2)^2 + 2 \lambda \sigma^2_a}{\sigma_a^2}} \), and formula (B.3) to determine its lower bound

\[ I_2 \geq -e^{-\gamma^2 T} \left[ N \left( \frac{-x - A_1 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_u}{a_0} \right)^{2A_1} N \left( \frac{-x + A_1 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right]. \quad (B.6) \]

Similarly, for the fourth term in the expected power utility function \( \alpha = x / \sigma_a \) and \( \beta = -A_4 \sigma_a \) we apply formula (B.3) to determine its upper bound

\[ I_4 \leq N \left( \frac{-x - A_4 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_u}{a_0} \right)^{A_1 - A_4} N \left( \frac{-x + A_4 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \]

and (B.4) to find the lower bound

\[ I_4 \geq e^{-\gamma^2 T} \left[ \left( \frac{K_u}{a_0} \right)^{A_1 + A_7} N \left( \frac{-x + A_7 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) + \left( \frac{K_u}{a_0} \right)^{A_1 - A_7} N \left( \frac{-x - A_7 \sigma^2_a T}{\sigma_a \sqrt{T}} \right) \right], \quad (B.8) \]

where \( A_7 = \sqrt{\frac{(\mu_a - \sigma^2_a / 2 + \gamma \rho \sigma_a \sigma_b)^2 + 2 \lambda \sigma^2_a}{\sigma_a^2}} \).

Finally, substituting (B.5) and (B.7) we obtain an upper bound of expected power utility:

\[ E_{\gamma} U \alpha = \frac{K^y}{A + \gamma} \left[ \int_{0}^{T} e^{-\gamma^2 t} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 \right) t}{\sigma_a \sqrt{T}} \right) dt \right] \]

\[ \quad - e^{-(\lambda + \gamma) \delta T} \int_{0}^{T} e^{-\gamma^2 t} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 \right) t}{\sigma_a \sqrt{T}} \right) dt \]

\[ + b_0^y e^{(\mu_a + (\gamma - 1)\sigma^2_a / 2 - \delta) T} \left[ \int_{T}^{\infty} \frac{|x|}{\sigma_a^2} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 + \gamma \rho \sigma_a \sigma_b \right) t}{\sigma_a \sqrt{T}} \right) dt \right] \]

\[ + e^{-\gamma^2 T} \int_{0}^{T} e^{-\gamma^2 t} \left( \frac{x - \left( \mu_a - \sigma^2_a / 2 + \gamma \rho \sigma_a \sigma_b \right) t}{\sigma_a \sqrt{T}} \right) dt \].
Using results (B.6) and (B.8) we obtain a lower bound of expected power utility:

\[ EG^U(\gamma) = K_u \frac{\lambda}{\lambda + \gamma \delta} \left[ \left( \frac{K_a}{a_0} \right)^{A_1 + A_2} + \frac{K_a}{a_0} \right] N \left( \frac{-x - A_2 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) + \left( K_a \frac{A_1 - A_2}{a_0} \right) N \left( \frac{-x + A_2 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) \]

\[ - e^{-\gamma T} \left[ N \left( \frac{-x - A_1 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) + \frac{K_a}{a_0} \left( \frac{x}{\sigma_a \sqrt{T}} \right)^{2A_1} \right] N \left( \frac{-x + A_1 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) \]

\[ + b^2 \left( \frac{x}{\sigma_a \sqrt{T}} \right)^{\gamma \mu_0 (1-10^2/2-\delta)-\delta T} \left[ N \left( \frac{-x - A_1 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) - \frac{K_a}{a_0} \left( \frac{x}{\sigma_a \sqrt{T}} \right)^{2A_1} \right] N \left( \frac{-x + A_1 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) \]

\[ + e^{-\gamma T} \left[ \left( \frac{K_a}{a_0} \right)^{A_1 + A_2} \right] N \left( \frac{-x + A_2 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) + \left( K_a \frac{A_1 - A_2}{a_0} \right) N \left( \frac{-x - A_2 \sigma_a^2 T}{\sigma_a \sqrt{T}} \right) \]

with \( A_1 = \frac{\mu_0 - \sigma_a^2 T}{\sigma_a^2}, A_2 = \frac{\sqrt{\mu_0 - \sigma_a^2 T^2 + 2 \mu_0 \sigma_a T}}{\sigma_a^2}, A_4 = \frac{\mu_0 - \sigma_a^2 T^2 + 2 \mu_0 \sigma_a T}{\sigma_a^2}, A_6 = \frac{\sqrt{\mu_0 - \sigma_a^2 T^2 + 2 \mu_0 \sigma_a T}}{\sigma_a^2}, A_7 = \frac{\sqrt{\mu_0 - \sigma_a^2 T^2 + 2 \mu_0 \sigma_a T}}{\sigma_a^2}. \)

References


